

# Case Study -Spark Fund Investment Analysis

By – Nanda Kishore Joshi

## Data Understanding

Our Understanding of Data:

Companies –This dataset has information about companies and their categories

Rounds –Dataset has information on investment preferences of the companies like amount, investment type, investment date etc.,

Mappings- Dataset maps the category list with the main sectors where he funds was invested

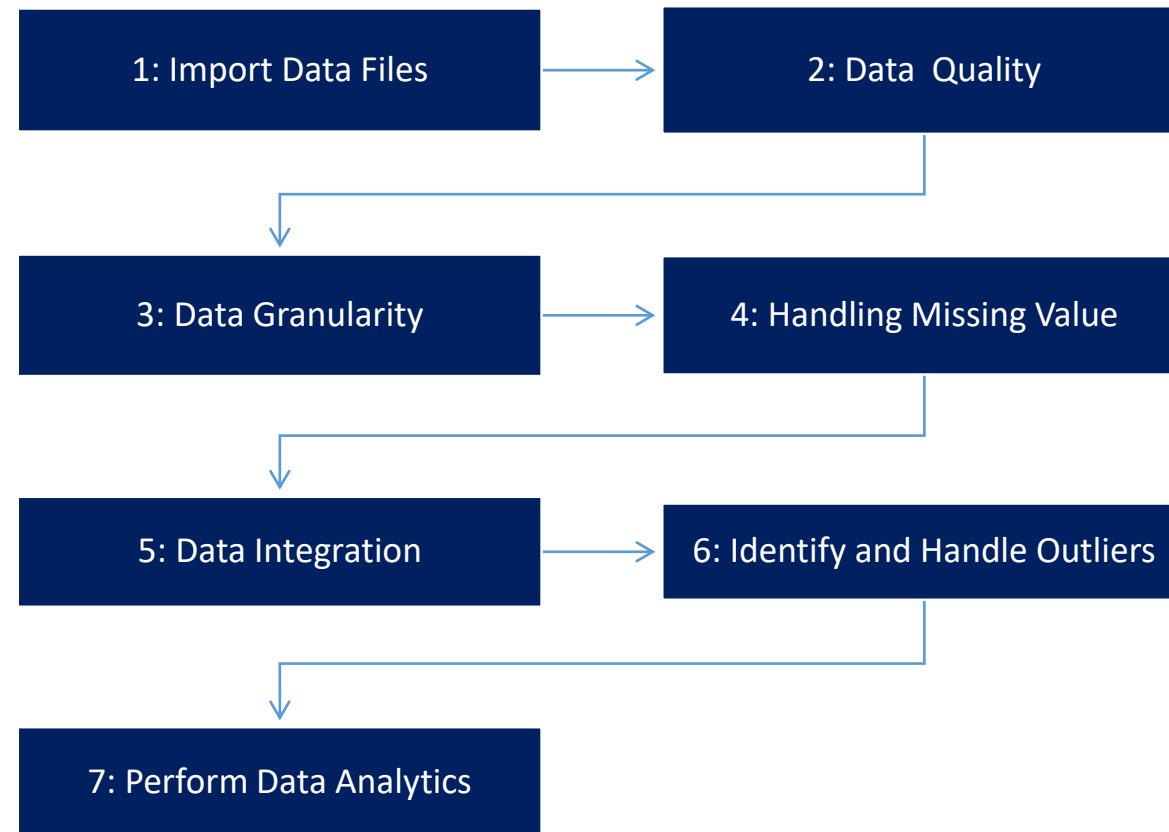
Key Stages/Objectives:

- ☐ Identify best Investment type.
- ☐ Identify best countries to Invest.
- ☐ Identify best Sectors to invest on.
- ☐ Identify best possible investment options.
- ☐ Plot graphical representation to showcase the findings.

# Our Approach -Spark Fund Investment Analysis

## Key Data Points

- Unique Companies -66368
- Permalink in companies file and company\_permalink in the round2 file are unique key
- There are 2 companies in the rounds2 file which are not present in companies
- There are approximately 114949 observation present in master file after merge
- Records with NaNs in Invested amount are removed. They constitute to 10% of the total records



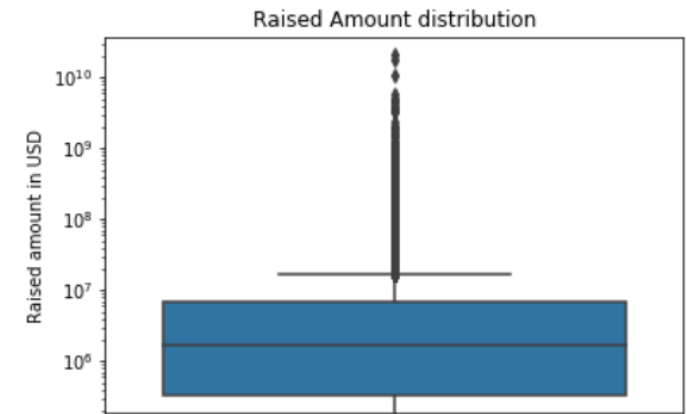
# Data Challenges Missing & Outliers

## Missing Data

- Raised amount is missing for 19990 records in rounds2 dataset
- Category List is missing for 3148 records in companies dataset
- 6958 records do not have country codes.

## Outliers/Wrong Data

- 415 companies have zero Raised amount
- Cambridge Technology Group has the highest raised amount of 10664943639 USD.
- Few records have wrong Category\_list in mapping dataset. '0' should be replaced by 'na' to correct the data.



## Data Redundancy

### Companies:

Below mentioned columns are redundant for the required analysis and can be removed from companies dataset

- Name
- Home\_url
- Status
- State\_Code
- Region
- City
- Founded at

### Rounds2 :

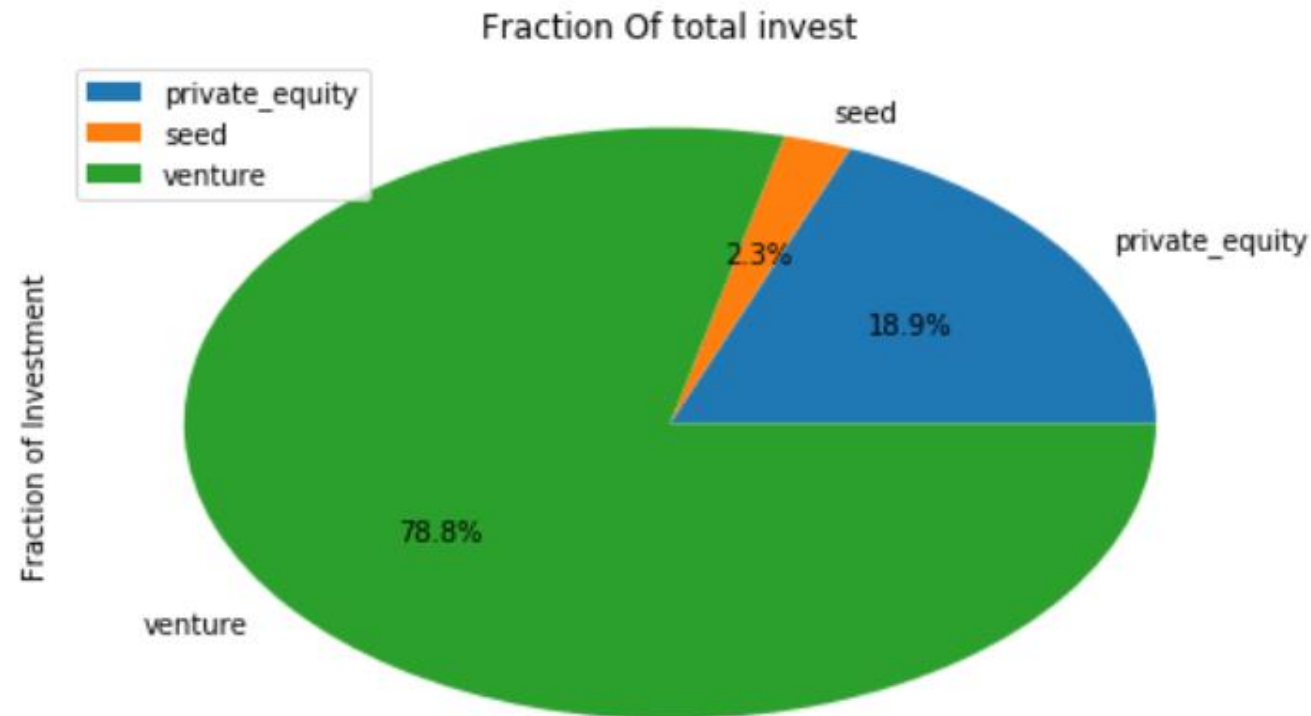
Below mentioned columns are redundant for the required analysis and can be removed from rounds2 dataset

- Funding\_round\_permalink
- funding\_round\_code
- funded\_at

# Graphical Representation of Data

Plot 1:

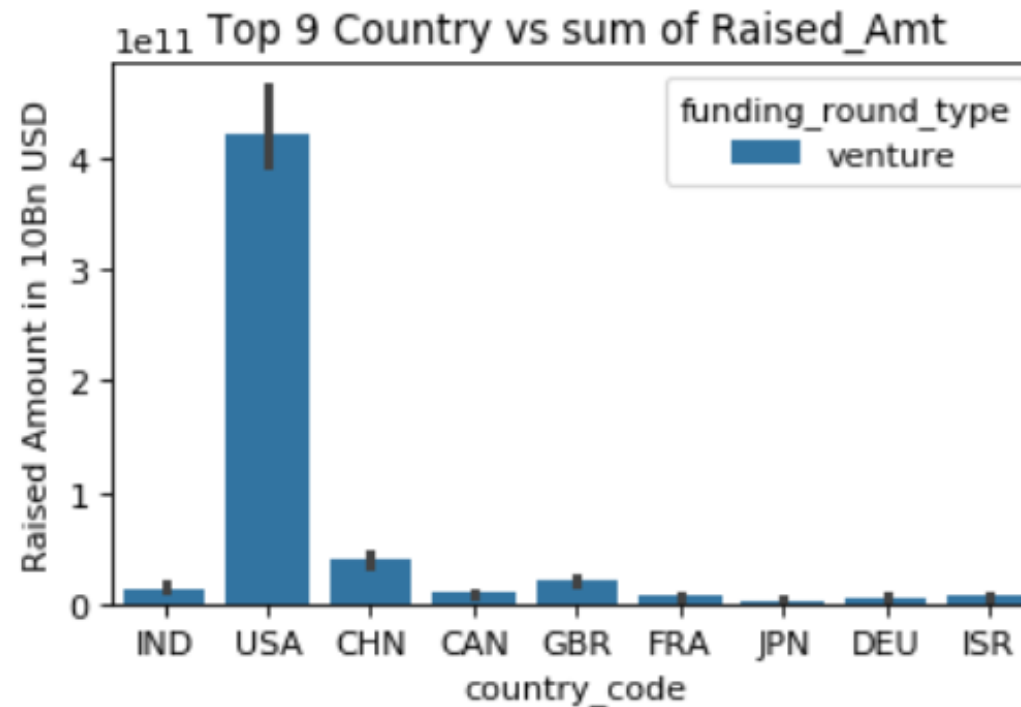
Plot showing the fraction of total investments (globally) in venture, seed, and private equity, and the average amount of investment in each funding type



# Graphical Representation of Data

Plot 2:

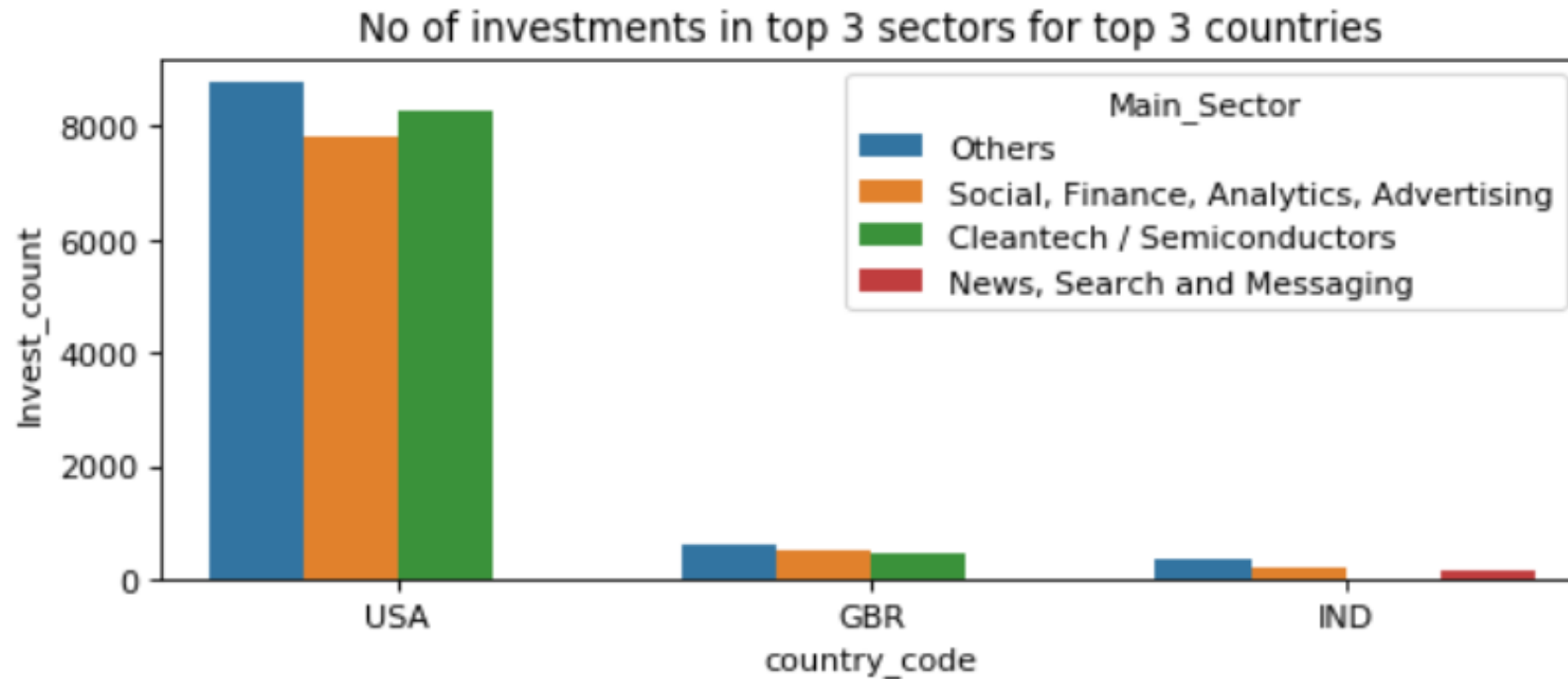
Plot showing the top 9 countries against the total amount of investments of funding type 'Venture'



# Graphical Representation of Data

Plot 3:

Plot showing the number of investments in the **top 3 sectors** of the **top 3 countries**





## Key Recommendation

- ☐ Best invest type in “Venture”
- ☐ Most suitable geography is United State of America
- ☐ Most suitable sector in USA is Others followed by social and Cleantech.
- ☐ Second most suitable geography is United Kingdom
- ☐ Most suitable sector in UK is Others followed by social and Cleantech.
- ☐ Third most suitable geography is India
- ☐ Most suitable sector in INDIA is Others followed by social and News.

Thank You